

# East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea)

Carlos Carreras<sup>1</sup> | Alex García-Cisneros<sup>2,3</sup> | Owen S. Wangensteen<sup>4</sup> |  
 Víctor Ordóñez<sup>1</sup> | Creu Palacín<sup>3</sup> | Marta Pascual<sup>1</sup> | Xavier Turon<sup>2</sup>

<sup>1</sup>Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Barcelona, Spain

<sup>2</sup>Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Girona, Spain

<sup>3</sup>Department of Evolutionary Biology, Ecology and Environmental Sciences and IRBio, University of Barcelona, Barcelona, Spain

<sup>4</sup>Norwegian College of Fishery Science, UiT The Arctic University of Norway, Tromsø, Norway

## Correspondence

Xavier Turon: Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala St. Francesc 14, 17300 Blanes, Girona, Spain.  
 Email: xturon@ceab.csic.es

## Funding information

Catalan Government, Grant/Award Number: SGR2017-1120; Spanish Government, Grant/Award Number: CTM2013-48163 and CTM2017-88080

Editor: Aibin Zhan

## Abstract

**Aim:** The Atlanto-Mediterranean edible purple sea urchin, *Paracentrotus lividus*, is a commercially exploited keystone species in benthic communities. Its browsing activity can deeply modify the littoral landscape, and changes in its abundance are of major conservation concern. This species is facing nowadays contrasting anthropogenic pressures linked to predator release, exploitation and sea warming. Management of this key species requires knowledge of its genetic structure, connectivity and local adaptation. Our goal was to assess the current global status of the species under a genomic perspective.

**Location:** Atlanto-Mediterranean shores from Morocco and France to Turkey.

**Methods:** We used genotyping by sequencing (GBS) of 241 individuals belonging to 11 populations spanning the known range of distribution of the species. We obtained 3,348 loci for population genomics and outlier analyses.

**Results:** We identified significant genetic structure and a gradient matching the longitudinal position of the localities. A hierarchical analysis revealed two main clusters (Atlantic and Mediterranean) and subtler patterns of differentiation within them. Candidate markers for selection identified between and within these two main clusters were mostly different, likely indicating different selective pressures. Adaptation to maximum salinity and maximum temperature appeared as an important driver of the transition between Atlantic and Mediterranean basins. Other stressors, such as minimum temperature or range of temperature, seem to define the structuring within the Mediterranean.

**Main conclusions:** Our study shows the potential of hierarchical analyses on population genomics to detect fine-scale genetic structure and adaptation signatures in marine species with long dispersal capabilities. Although genetic interchange occurs

Carlos Carreras and Alex García-Cisneros contributed equally as first authors.

Marta Pascual and Xavier Turon contributed equally as senior authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd

widely in *Paracentrotus lividus*, the species is sensitive to dispersal barriers, displays isolation by distance and faces local selective pressures associated to environmental conditions, all of which can render it more vulnerable than previously thought.

#### KEYWORDS

benthic communities, engineer species, genotyping by sequencing, oceanographic discontinuities, outlier analysis, population genomics, sea urchin

## 1 | INTRODUCTION

The population structuring of living organisms is shaped by the combination of different evolutionary forces such as migration, genetic drift or selection. Historically, the selection component has been underestimated in relation to other processes in population genetic analyses. However, the rise of genomic studies during the last decade has opened new windows for research including the analysis of outlier loci to look for genomic signals of local adaptation or associated to environmental cues (Benestan, Ferchaud, et al., 2016). The genotyping of thousands of markers spanning the whole genome has increased the power to detect even subtle genetic patterns previously unnoticed using markers such as microsatellites or mitochondrial DNA sequences (Bradbury et al., 2015, 2010). Furthermore, the analysis of outlier loci has revealed additional levels of structuring in many cases (Milano et al., 2014; Vandamme et al., 2014) showing trade-offs between selection and dispersal (Carreras et al., 2017). This additional improvement of population genomics, when compared to conventional population genetics, is especially relevant in marine organisms, as they include widely dispersing species and feature large population sizes rendering genetic differentiation across populations difficult to detect (Lamichhaney et al., 2012; Xuereb et al., 2018). For this reason, the inclusion of adaptive genetic variation is necessary to identify population structuring mediated by local adaptation to environmental conditions (Sandoval-Castillo, Robinson, Hart, Strain, & Beheregaray, 2018). The use of genome-wide markers increases the chance of identifying regions under selection and has been recommended for delineating conservation and management units of natural populations (Funk, McKay, Hohenlohe, & Allendorf, 2012).

The assessment of genomic signals of selection on non-model organisms is usually performed in two different ways (Ahrens et al., 2018). On one hand, outlier analyses (OAs) are based on the search of genome-wide markers showing outlier values of genetic differentiation ( $F_{ST}$ ) among the studied populations as they are considered to be candidate loci for local adaptation (Carreras et al., 2017). However, the power to find such outlier loci depends, by definition, on the mean values of genetic differentiation found among populations. Thus, if different levels of genetic differentiation exist (e.g. a deep genetic break separating two groups of genetically differentiated populations) the power to detect outliers and signals of local adaptation at the lower levels (i.e. within groups) may be compromised and obscured by patterns occurring between genetic groups. On the

other hand, environmental association analyses (EAAs) use environmental variables to find genomic signatures associated to them, and thus potentially under selection (Benestan, Quinn, et al., 2016). Again, these environmental variables can show different patterns in the different genetic groups (that likely occur in different geographic zones), and thus, only the associations related to the major breaks may be revealed. If the overall genetic structure is not accounted for in the assessment of genomic signals of selection, the results are likely to be partial. Under these circumstances, a hierarchical approach combining the analyses at different levels is fundamental to get out the most of genetic data (Carreras-Carbonell, Macpherson, & Pascual, 2006). Although several constraints have been described for the analysis of candidate loci under selection using either OAs or EAAs (Ahrens et al., 2018), the impact of hierarchical genetic structuring on the search of selection signatures at different scales has not been addressed, and this may be a critical issue, especially on species with wide distribution ranges.

In this study, we applied a genomics approach to assess the population structure of the edible purple sea urchin *Paracentrotus lividus* (Lamarck, 1816). This species has a wide distribution range across the Mediterranean Sea and north-east Atlantic. This sea urchin is a key component in benthic communities due to its grazing activity that regulates the density of macrophyte populations (Agnetta et al., 2015; Benedetti-Cecchi, Bulleri, & Cinelli, 1998; Palacín, Giribet, Garner, Dantart, & Turon, 1998; Sala, Boudouresque, & Harmelin-Vivien, 1998; Wangensteen et al., 2011). Changes in its abundance are therefore of major concern for the conservation of littoral communities. This species is nowadays facing contrasting anthropogenic pressures. While overfishing of predators can promote high densities of sea urchins (Hereu, Zabala, Linares, & Sala, 2005; Sala et al., 1998), this species is also exploited for their roe (Barnes & Crook, 2001), and it has shown early signs of collapse in some areas. This decline is probably the result of the combined impact of over-exploitation, climate change due to its temperature tolerances (Privitera, Noli, Falugi, & Chiantore, 2011; Yeruham, Rilov, Shpigel, & Abelson, 2015) and possible competition with other species (Rilov, 2016; Yeruham et al., 2015). A sound knowledge of the genetic structure, connectivity and local adaptation of this key species is necessary for a proper management.

The adults of *P. lividus* are known to have limited mobility (Palacín, Giribet, & Turon, 1997) but their long pelagic larval duration (PLD) of 20–40 days (Pedrotti, 1993), heterogeneity in settlement (Hereu, Zabala, Linares, & Sala, 2004; Tomas, Romero, & Turon,

**TABLE 1** Sampling localities for *Paracentrotus lividus* including the code, name, geographic coordinates, date of sampling, number of sampled individuals (N), mean number of reads per individual, mean number of alleles per locus, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities, inbreeding coefficient ( $F_{is}$ ), number of monomorphic loci (M) and number of loci out of Hardy-Weinberg equilibrium after the FDR correction (HWE)

Code	Sampling locality	Coordinates	Date	N	Reads $\times 10^6$ (SD)	Mean alleles/ locus (SD)	$H_o$	$H_e$	$F_{is}$	M	HWE
CZ	Crozon	48.207; -4.377	December 2014	23	6.23 (1.68)	7.55 (5.14)	0.365	0.496	0.263	152	553
SI	Sidi Ifni	29.391; -10.176	October 2014	22	3.09 (1.08)	7.38 (4.97)	0.367	0.504	0.271	134	541
TA	Tarifa	36.000; -5.609	September 2014	22	4.29 (2.00)	7.57 (4.82)	0.379	0.511	0.257	114	564
LH	La Herradura	36.721; -3.728	September 2014	22	3.40 (1.36)	7.53 (4.74)	0.378	0.509	0.257	109	568
CA	Carboneras	36.992; -1.886	September 2014	22	5.06 (1.42)	7.48 (4.73)	0.369	0.498	0.258	117	578
PA	Palos	37.635; -0.689	September 2014	22	3.60 (1.63)	7.30 (4.65)	0.370	0.494	0.252	122	507
XA	Xàbia	38.753; 0.2291	September 2014	22	5.49 (1.18)	7.38 (4.71)	0.371	0.489	0.242	118	545
CO	Colera	42.391; 3.155	February 2015	22	5.48 (2.68)	7.34 (4.66)	0.369	0.491	0.249	108	531
TR	Trieste	45.691; 13.741	December 2014	22	4.69 (2.18)	7.24 (4.72)	0.367	0.485	0.242	132	494
FI	Fier	40.662; 19.280	May 2015	22	4.17 (1.07)	7.22 (4.71)	0.366	0.483	0.241	126	516
DL	Dalyan	36.764; 28.613	April 2015	20	5.41 (2.84)	6.73 (4.32)	0.368	0.482	0.237	132	442

Note: All  $F_{is}$  values were significant after FDR correction. The total number of loci is 3,348 in all localities with the exception of Sidi Ifni (3,343 loci) and Crozon (3,337 loci). Average values include standard deviation in parentheses (SD).

2004) and broadcast spawning behaviour (Minchin, 1992) suggest that this species may have low genetic differentiation among localities. Previous studies assessing the population structure of *P. lividus* using different sets of mitochondrial and nuclear DNA markers found significant genetic differentiation only across two known oceanic discontinuities, the Almería-Oran front separating the Atlantic and the Mediterranean populations (Calderón, Giribet, & Turon, 2008; Duran, Palacín, Becerro, Turon, & Giribet, 2004), and the Adriatic-Ionian front (Maltagliati, Giuseppe, Barbieri, Castelli, & Dini, 2010). Temporal variability among different cohorts of the same localities was also assessed suggesting little temporal variation, with some exceptions associated with abnormalities of circulation patterns (Calderón, Palacín, Palacín, & Turon, 2009; Calderón, Pita, Brusciotti, Palacín, & Turon, 2012) or chaotic genetic patchiness (Couvray & Coupé, 2018). However, it has been suggested that *P. lividus* might present higher levels of structuring across the Atlantic and Mediterranean than previously suggested (Penant, Aurelle, Feral, & Chenuil, 2013). This potential structuring at finer scale was evaluated within the Adriatic Sea with a population genomics approach using 2b-RAD coupled with Lagrangian simulations (Paterno et al., 2017). This first genomic study showed differentiation between the Adriatic-Ionian populations and two external populations (France and Tunisia), although no fine-scale structuring was found within the Adriatic. Furthermore, this study did not provide a general picture for *P. lividus*, given that its spatial scope did not cover the wide range of the species' distribution.

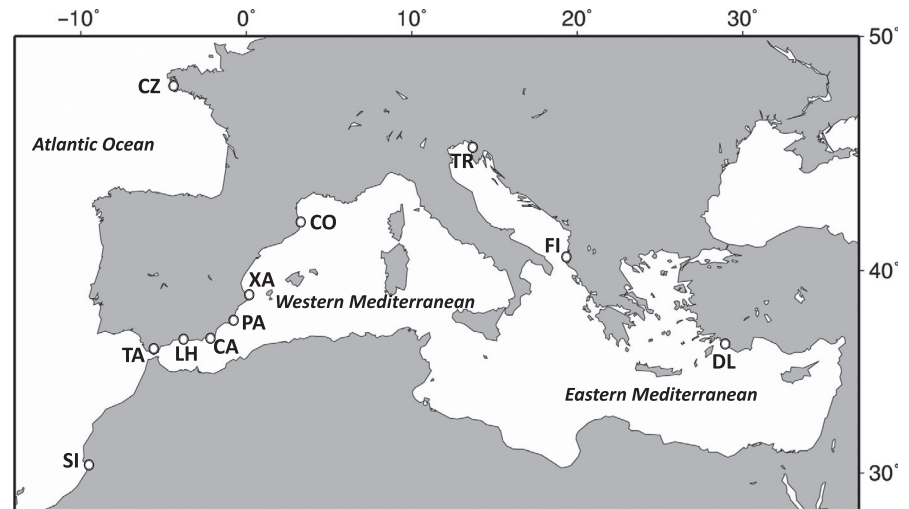
In this context, we used genotyping by sequencing (GBS) on 241 individuals of *P. lividus* sampled from 11 localities spanning its Atlanto-Mediterranean distribution range. The specific aims of this study are (a) to determine the global population structure of this key-stone species using a panel of genome-wide markers, (b) to test the existence of selective pressures by identifying candidate loci under selection between and within genetic groups, (c) to relate the signal of these candidate loci to environmental variables and (d) to ascertain adaptation drivers by relating candidate loci under selection to known functions through gene annotation. Finally, we discuss the implications of our results in the context of management and conservation considering the species' current exploitation and foreseeable population regression due to environmental alteration and the complexity of its distribution range.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling

A total of 241 *P. lividus* individuals were sampled by SCUBA diving, snorkelling or by hand during low tide within a period of 17 months (from June 2014 to October 2015) from 11 localities (between 20 and 23 individuals per locality, Table 1; Figure 1). The sampling localities are distributed in three major regions: (a) two sites in the eastern Atlantic, Crozon (CZ, France) and Sidi Ifni (SI, Morocco); (b) six sites in the western Mediterranean, Tarifa (TA), La Herradura

**FIGURE 1** Map of the sampling localities with their code names. CZ: Crozon; SI: Sidi Ifni; TA: Tarifa; LH: La Herradura; CA: Carboneras; PA: Palos; XA: Xàbia; CO: Colera; TR: Trieste; FI: Fier and DL: Dalyan. Map created using the free software MAPTOOL (SEATURTLE.ORG Maptool. 2002. SEATURTLE.ORG, Inc. <http://www.seaturtle.org/maptool/>, February 26, 2018)



(LH), Carboneras (CA), Palos (PA), Xàbia (XA) and Colera (CO); and (c) three sites in the eastern Mediterranean, Trieste (TR, Italy), Fier (FI, Albania) and Dalyan (DL, Turkey). The sampled localities span c. 5,000 km (shortest distance by sea) from CZ to DL. Samples were stored in absolute ethanol, and between 10 and 15 mg of muscular tissue from the Aristotle's lantern from each individual was used for DNA extraction with the QIAMP® DNA Mini Kit (QIAGEN) following the manufacturer's instructions.

## 2.2 | Genotyping by sequencing and loci selection

A GBS protocol (Elshire et al., 2011) was carried out at the "Centre Nacional d'Anàlisi Genòmica" (CNAG). The genome was digested using the restriction enzyme ApeKI (selected based on fragment size distribution in a preliminary assay), and the resulting fragments were ligated to individual barcodes and to common adapters. DNA fragments were paired-end sequenced ( $2 \times 100$  bp) in an Illumina HiSeq 2000 platform using 96 multiplexed individuals in each lane of a flow cell. Each individual was sequenced two or three times in order to increase sequence depth, totalling 10 lanes for this study.

The raw Illumina sequences were analysed using the GlibPSs toolkit, designed to analyse data from massive sequencing of non-model organisms (Hapke & Thiele, 2016). GlibPSs can handle paired-end sequences; it does not require the same sequence length at different loci and implements an indel detection method to flag potentially problematic loci that would include alleles of different sizes. All sequences were trimmed to 80 bp to discard the lower quality bases at the ends, and the 5' restriction sites were removed. The default values for filtering were modified to increase stringency, and thus, sequences shorter than 32 bp, with N's, or with a phred score quality lower than 22 in a sliding window of 5 bp, were discarded. A minimum overlap of 5 bp was required to assemble forward and reverse sequences of a paired-end read, and thus, the maximum length of a locus after assembling both paired-end sequences was 152 bp. Staggered read pair problems arose with loci shorter than the read length because the two restriction sites were very close.

In these cases, forward and reverse reads corresponded to exactly the same fragment including the restriction sequence at both ends. Depending on the length of the fragment, the initial 80 bp trimming may not completely cut off the restriction enzyme sequence at the 3' end of each read, while the restriction sites at the 5' ends were removed. In these cases, the program could not overlap the sequences and appended them. Those duplicate sequences were identified, and only the forward read was kept for further analyses.

Locus search with GlibPSs was first carried out separately for each individual. Identical sequences in a given individual were grouped in the same cluster (*svar* in the terminology of the program). Different *svars* were grouped for each individual into loci using the default distance settings. A minimum depth of at least 5 sequences was necessary for a locus to be deemed as valid. We defined as alleles of individual genotypes those *svars* representing more than 30% of the total number of reads in a locus. Thus, heterozygote individuals were those having two alleles above this threshold as genotypes based on an uneven distribution of alleles were removed. Loci were identified as haplotype loci, based on the procedure used by GlibPSs and other software for "de novo" loci identification. Thus, all the polymorphic sites of the different sequences (*svars*) of each locus were included for identifying alleles. We used haplotype loci as previous studies have shown that working with sequence variants (multiallelic) instead of using only one SNP per locus (biallelic) is more informative (Casso, Turon, & Pascual, 2019; Ryman et al., 2006). A global loci catalogue was assembled with all the alleles found in all the individuals using the default distance settings. Loci flagged by the program as potentially including an indel or with more than two alleles per individual were discarded, as they can be the result of combining paralogous loci, as well as PCR or sequencing errors. We also removed deeply sequenced loci, as they are likely paralogous DNA regions with multiple copies along the genome. Following the algorithm developed in GlibPSs, a median depth percentile below 0.2, as scaled across individuals, was used to identify and remove these deeply sequenced loci. A final locus selection was performed to retain only loci present in 70% of the total number of individuals and to obtain the haplotype loci dataset. Nonetheless, for comparison with our

haplotype loci dataset and for the requirements to use SNPs in some programs (see below) we also selected from each retained haplotype locus only the first SNP, using the option `-uS 2` of GlibPSs, to obtain a new dataset, henceforth called the SNP loci dataset.

## 2.3 | Population structure

Hardy–Weinberg equilibrium (HWE) was assessed for each locus in each sampled locality with the “pegas” R package (Paradis, 2010). Significant departures from HWE were detected after a Benjamini–Yekutieli (B-Y) FDR correction for multiple comparisons (White, van der Ende, & Nichols, 2019). Markers with significant departures from HWE in 60% or more localities were removed from the final dataset, to avoid using the combination of paralogous loci in the same locus as they are expected to be consistently out of HWE across populations (Benestan et al., 2015).

Pairwise differences between localities were calculated with Nei's estimator of  $F_{ST}$  (Nei, 1987) in the “hierfstat” R package, using 999 permutations to calculate the respective  $p$ -values (Goudet, 2005) with the B-Y FDR correction. The results were represented with heatmaps and dendrograms with the “gplots” R package (Warnes et al., 2016). Observed and expected heterozygosities, allelic richness and  $F_{IS}$  values per sampling location were calculated using the R packages “genetics” and “hierfstat” (Goudet, 2005; Warnes, Gorjanc, Leisch, & Man, 2019). In order to identify isolation by distance (IBD), a Mantel test was performed with the “vegan” R package (Dixon, 2003; Oksanen et al., 2019) using the shortest distance by sea between localities measured with Google Earth.

A discriminant analysis of principal components (DAPC) was performed with the “adegenet” R package (Jombart, 2008). We retained a total of  $N/3$  PCs, being  $N$  the number of individuals, and 3 discriminant functions using localities as prior grouping information. Additionally, we used the Bayesian information criterion (BIC) to detect the optimal number of genetic clusters. We performed a  $k$ -means clustering from  $k = 1$  to  $k = 10$ , using  $10^5$  iterations, and, in this case, we retained 250 PCs as the cumulative variance explained by the eigenvalues of the PCA was saturated at this value. The results obtained were used to separate our samples per cluster that were represented in new DAPCs to better identify the genetic structure within each genetic cluster. Again, we retained a total of  $N/3$  PCs and 3 discriminant functions using localities as prior grouping information. Additionally, population structuring was also evaluated using STRUCTURE, which implements a Bayesian clustering method to identify the most likely number of populations ( $K$ ). We used the strategy and parameters described in the literature (Evanno, Regnaut, & Goudet, 2005), and thus, we carried out 10 runs per each value of  $K$  ranging from 1 to 10. We used the model of correlated allele frequencies and a burn-in of 50,000 followed by 500,000 Markov Chain Monte Carlo steps. We estimated the ad hoc statistic  $\Delta K$  in order to infer the most likely number of populations using STRUCTURE HARVESTER (Earl & vonHoldt, 2012). The 10 runs of STRUCTURE for the most probable  $K$  were averaged using

CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007). To test the resolution of detecting the population genetic structure of the SNP loci dataset compared to the haplotype loci dataset, we also carried with the SNP loci dataset a DAPC analysis, computed pairwise  $F_{ST}$  values among localities and performed a STRUCTURE analysis to identify the number of genetic groups as described above.

## 2.4 | Identification of loci candidate for local adaptation

Four different sample groupings were used to detect candidate loci among the haplotype loci dataset based on outlier  $F_{ST}$  (OAs): “All populations,” “Atlantic versus Mediterranean,” “Atlantic” and “Mediterranean.” In the “All populations” grouping, all samples were included and grouped by sampling locality to obtain a general overview of the outlier loci across the whole geographical range sampled. In the “Atlantic versus Mediterranean” grouping, all the samples were also included but they were combined in the two groups as identified by the  $k$ -means method in DAPC belonging to the Mediterranean and the Atlantic clusters (see Section 3). This analysis was set in order to detect outlier loci differentiating the Atlanto-Mediterranean transition. Finally, each cluster was analysed independently with their samples grouped by locality in order to explore the outliers of the internal structuring within each group (“Atlantic” or “Mediterranean”).

In the four groupings, outliers were identified with Arlequin (Excoffier & Lischer, 2010), Lositan (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) and BayeScan (Foll & Gaggiotti, 2008). The analyses with Arlequin were performed using 100,000 coalescent simulations and 100 demes in a finite Island model. Lositan analyses were performed with 1,000,000 simulations and a subsample of 50, in an infinite allele model, using a “neutral” mean  $F_{ST}$  (which means that the software performs a preliminary simulation to remove potentially selected loci before computing the mean neutral  $F_{ST}$ ). Benjamini–Yekutieli FDR corrections (White et al., 2019) were applied to the  $p$ -values of Arlequin and Lositan with an overall alpha level of 0.01. BayeScan search was implemented with a total of 150,000 iterations, with a burn-in period of 100,000 and 40 pilot runs of 10,000 iterations each. Outlier loci in BayeScan were revealed using a  $q$ -value of 0.01. Only loci detected simultaneously with the two more conservative programs (see Section 3) were considered as outliers in order to minimize the risk of obtaining false positives.

The same analyses of population structure performed with all loci were repeated using only the outlier loci. Thus, DAPC representations and  $F_{ST}$  comparisons were performed separately for the three sets of outlier loci as indicated above by using “All populations,” “Atlantic versus Mediterranean” or only “Mediterranean” samples. No analyses were performed within the Atlantic cluster as only two outlier loci were detected (see Section 3).

The  $F_{ST}$  matrices of all subsets of outliers were compared to environmental variables that could explain the genetic structuring found. Both monthly mean sea surface temperature and salinity data were



obtained from Copernicus Marine Environment Monitoring Service database (CMEMS, <http://marine.copernicus.eu>, product identifier: MULTIOBS\_GLO\_PHY\_REP\_015\_002) using the R package “ncdf4” version 1.16 (Pierce, 2017) from January 1993 to December 2016 (288 months) and averaged. For each abiotic variable, we computed distance matrices using the pairwise differences in the mean, maximum, minimum and range of the variable. These distance matrices were compared individually with the  $F_{ST}$  matrices using partial Mantel tests as implemented in the R package “vegan” (Dixon, 2003; Oksanen et al., 2019) and using the geographic distance among populations as a controlling variable.

We used a redundancy analysis (RDA) (Forester, Lasky, Wagner, & Urban, 2018) as an EAA in order to identify candidate loci significantly associated to environmental variables (association loci). Prior to perform the RDA, we imputed missing genotypes on the SNP loci dataset by replacing them with the most common genotype across all individuals. Finally, we identified significantly associated loci using as response variable the matrix of SNP genotypes of all loci and the environmental variables as predictors. The RDA was performed with all the localities and with only the Mediterranean localities but excluding the transition zone.

## 2.5 | Regionally missing loci

We looked for geographic patterns with regionally missing loci, defined as the haplotype loci that did not pass the <30% missingness filter and were therefore discarded from the final dataset. For each one of the four identified areas (Atlantic, transition zone, western Mediterranean and eastern Mediterranean, see Section 3, Figure 2), we performed independent analysis using the non-filtered loci dataset. For each one of these partial datasets, we applied the same filters used for the whole dataset, as some of the filters may be dependent on the samples analysed. For all loci present in at least one of the partial datasets but discarded in the whole dataset, we estimated their frequency (percentage of individuals with these loci per population). We plotted their frequency across all localities using a heatmap and dendrogram with the “gplots” R package (Warnes et al., 2016).

## 2.6 | BLAST of candidate loci for local adaptation

The sequences of the candidate loci for local adaptation, found with either OAs (outlier loci) or EAAs (association loci), were blasted against the genome of *Strongylocentrotus purpuratus* in the NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) using the blastn option. We also blasted the sequences of all the regional missing loci that were completely absent in some of the partial datasets. The scaffold accession number, nucleotide position, name of the gene where the sequence is located (if applicable) and position (exon, intron or intergenic) were recorded for matches with  $E$ -values below  $1 \times 10^{-4}$ . Sequences with a

positive blastn hit (i.e. below the  $E$ -value threshold) were analysed using Blast2GO 4.1.5 software (Gotz et al., 2008). First, a blastx algorithm was applied using the NCBI Blast service with a taxonomy filter for “sea urchin” (taxid 7625) and the same  $E$ -value of  $1 \times 10^{-4}$  used for the blastn analysis. Positive matches were mapped and annotated using the default parameters of the program. Sequences with a blastn hit for which an ontology term could not be found were manually checked using UniProt to identify potential functions as described for other echinoderms. We used the GO terms of the annotated genes to produce treemaps of the biological samples using REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011).

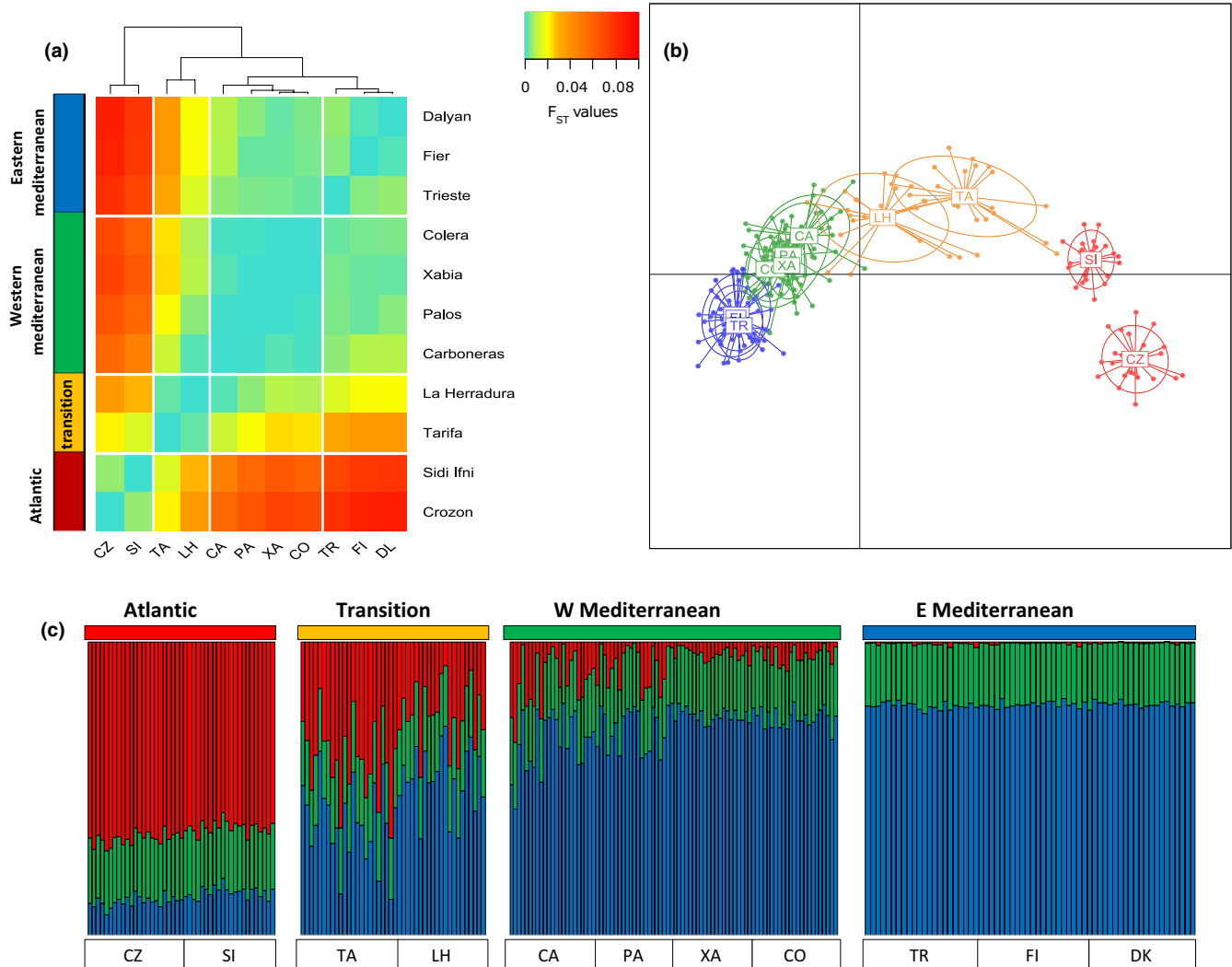
## 3 | RESULTS

### 3.1 | Locus identification

For the 241 assayed specimens, 4,613,169 reads were obtained on average per individual (minimum 800,000 reads, Table 1), and a mean of 79% were retained after applying the quality filter of the GibPSs—fdm software. These reads resulted in a total of 8,345,562 non-filtered loci across all samples. We then selected the loci with a maximum of 2 alleles per individual and removed potential indel variants. When we selected loci shared by at least 170 individuals (70% of all individuals), a total of 3,862 loci were retained. We applied additional filtering by removing deeply sequenced loci and loci in Hardy–Weinberg disequilibrium in at least 7 (c. 60%) of the localities to obtain a final dataset of 3,348 haplotype loci for further analyses.

### 3.2 | Population genetic diversity

The selected loci had different lengths, from 32 to 152 bp ( $123.6 \pm 34.3$ , mean  $\pm$  SD), and the number of polymorphic positions in each locus was variable ( $25.9 \pm 13.7$ ). Therefore, the total number of alleles per locus had a wide range, with 64 loci having more than 100 alleles and 9 loci having only 2 alleles (Figure S1). The mean number of alleles per locus was  $31.9 \pm 21.8$ . However, the mean number of alleles found within a given locality was more homogeneous, with an average of  $7.3 \pm 0.2$  alleles per locus (Table 1). Heterozygosity values in all localities were remarkably similar ( $H_o = 0.370 \pm 0.005$  and  $H_e = 0.494 \pm 0.010$ ). Even after removal of loci out of HWE equilibrium in the filtering step,  $F_{IS}$  values were positive and significant ( $0.252 \pm 0.011$ ) in all localities (Table 1) because an average of  $530.8 \pm 39.4$  retained loci was out of HWE per locality. Most of the loci were polymorphic within each locality; only  $3.7 \pm 0.4\%$  of loci were monomorphic at a given sampling locality (Table 1). Interestingly, a total of 5 loci were missing in all individuals of the two Atlantic localities, and an additional 6 loci were missing in all individuals from Crozon (French Atlantic). All other markers were found in at least some individuals from all localities.



**FIGURE 2** Population structuring analyses of *Paracentrotus lividus* from Mediterranean and Atlantic sites with the 3,348 haplotype loci of the global dataset. (a) Heatmap and dendrogram based on  $F_{ST}$  pairwise distances. (b) Discriminant Analysis of Principal Components (DAPC) with localities as prior grouping information. (c) Posterior probabilities of individual assignments to the most probable number of clusters using STRUCTURE ( $K = 3$ ). Sampling sites are as follows: CZ: Crozon; SI: Sidi Ifni; TA: Tarifa; LH: La Herradura; CA: Carboneras; PA: Palos; XA: Xàbia; CO: Colera; TR: Trieste, FI: Fier and DL: Dalyan

### 3.3 | Population structure and connectivity

$F_{ST}$  pairwise distances were low but significant after FDR correction in 37 out of 56 comparisons (Figure 2a; Table S1). The two Atlantic localities (CZ and SI) presented the highest  $F_{ST}$  values with all other localities (significant in all cases), but the  $F_{ST}$  value between them was much lower although significant. Furthermore, we also detected differences among the Mediterranean localities. TA and LH, two sites located in the Alboran Sea (Figure 1), were not significantly differentiated from each other, but differed from all other Mediterranean localities (with the exception of the comparison LH-CA). The three easternmost Mediterranean localities (TR, FI and DL) were significantly differentiated from the three westernmost ones (TA, LH and CA). The two Adriatic populations, TR and FI, were also significantly differentiated from each other. The clustering dendrogram using the  $F_{ST}$  pairwise matrix showed an Atlantic cluster, a transition cluster

(comprising the Alboran Sea localities) and a Mediterranean cluster further subdivided in eastern and western Mediterranean sites (Figure 2a). Similarly, the DAPC analysis showed a clear arch effect following a longitudinal geographic gradient of genetic differentiation, with Atlantic localities (CZ and SI) plotted at one extreme, the three localities from the eastern Mediterranean (TR, FI and DL) at the other extreme and all the other localities in-between sorted according to their longitudinal geographic order (Figure 2b).

Two clusters were identified following the Bayesian information criterion (BIC) that could be associated to Atlantic and Mediterranean affinities (Figure S2a; Table S2). All the individuals from the Atlantic localities were assigned to one cluster. The localities from the Alboran Sea (TA and LH) had a mixed composition, with a decreasing abundance of Atlantic-type individuals as we move eastwards following the coastline, which is consistent with the idea that this is a transitional area. All other localities were made up exclusively of

individuals of the second cluster. Therefore, we will refer to all the individuals of cluster 1 as the Atlantic cluster and all the individuals of cluster 2 as the Mediterranean cluster (Table S2). When analysing only the individuals of the Atlantic cluster with a DAPC representation, both Atlantic sites (CZ, SI) were separated while Tarifa and La Herradura clearly overlapped (Figure S3a). On the other hand, when representing the individuals of the Mediterranean cluster, the three localities from eastern Mediterranean (TR, FI and DL) were clearly separated from the others along the first axis (Figure S3b), while the Alboran Sea populations (TA and LH) lay at the other extreme of this axis. The Bayesian approach of STRUCTURE suggested the existence of three genetic groups ( $K = 3$ , Figure S2b). The distribution of the posterior probabilities showed again the existence of differences between the Atlantic and Mediterranean individuals with a transition zone (Figure 2c). The set of SNP loci used for comparison yielded overall similar population structuring than the haplotype loci dataset with an arch effect of genetic differentiation following the longitudinal geographic gradient (Figure 2 compared to Figure S4). However, no definition of the transition zone was observed in the heatmap based on  $F_{ST}$  values, and the differentiation between eastern and western Mediterranean localities was not as clear-cut based on the DAPC analysis, as it presented a higher overlap (Figure S4). Moreover, the number of clusters inferred by STRUCTURE ( $K = 4$ ) was not clearly observed in the assigned individual probabilities (Figure S4).

Considering all localities and the haplotype loci dataset, genetic differentiation significantly increased with geographic distance (Mantel test,  $R^2 = .618$ ,  $p = .002$ ). Furthermore, this relationship was significant, albeit with a lower correlation coefficient ( $R^2 = .514$ ,  $p = .003$ ) when only Mediterranean localities were considered, indicating that the significant isolation by distance detected is not only due to the genetic distances separating the Atlantic and the Mediterranean populations. Furthermore, the Mantel tests plots indicated that these results are unlikely to be caused by the presence of two differentiated groups but to a progressive isolation by distance (Figure S5).

### 3.4 | Signals of selection

The number of outlier loci detected by OAs was larger when comparing all samples grouped by locality (Table 2). Arlequin software and Lositan software were more conservative and generally detected fewer outlier markers than BayeScan, and thus, they were used for outlier selection in most searches (Table 2). After retaining the loci of the two more conservative programs, we did not detect any outlier locus presenting lower  $F_{ST}$  than expected under neutrality, and thus possibly under balancing selection. Furthermore, the different groupings not only revealed different number of outliers but also the outliers found were different in many cases (Figure 3). Most outlier loci found using all the samples grouped by locality ("All populations" outlier dataset) were also detected when comparing the Mediterranean and Atlantic clusters. Nevertheless, only c.

50% of outlier loci found within the Mediterranean and within the Atlantic clusters were also found in the "All populations" outlier dataset. The outlier loci identified when comparing the two clusters were mostly different to those identified within clusters (Supplementary Dataset). Mean global  $F_{ST}$  value per locus was always larger when using outlier than no outlier loci for any the three different programs used (Arlequin, Bayescan and Lositan) and by the four different outlier analysis performed (Figure S6).

As expected, the range and the mean values of pairwise genetic distances were larger using any set of outlier markers than using all loci (Table 2). Pairwise genetic differences using all loci (Figure 4a) and the three sets of outlier loci (Figure 4b-d) showed clear and significant genetic differentiation between Atlantic and Mediterranean localities with the exception of SI and TA for two sets of outlier loci (Tables S3–S5). The two sampling sites located in the Alboran Sea (TA and LH), the transition area between the Atlantic and Mediterranean, had significant differences with most other sites when "All populations" or "Atlantic versus Mediterranean" outliers were used (Figure 4b,c; Tables S3–S4). However, when the 31 Mediterranean outliers were used, genetic differentiation was not found among the transition area and most western Mediterranean sites (Table S5). Nonetheless, this set of outliers showed a marked differentiation between western and eastern Mediterranean localities, as well as within the latter (Figure 4d; Table S5). DAPC representations obtained with the different sets of loci (Figure S7) were basically similar to those obtained with all loci (Figure 2 and Figure S3), albeit the locality closer to the Gibraltar Strait (TA) appears separated in the representation of the Mediterranean cluster with the outliers found in the groupings of "All localities" and "Atlantic versus Mediterranean." The 31 outliers found in the Mediterranean grouping showed overall less resolution, but allowed a clearer and significant differentiation among the Turkish and Adriatic localities as well as between the eastern and western Mediterranean localities (Figure S7; Table S5).

Several pairwise population distances for several environmental variables were significantly correlated to  $F_{ST}$  matrices based on outlier loci after controlling for the effect of geographic distance with partial Mantel tests (Table S6). The genetic distances calculated using the outliers from all localities significantly correlated to maximum temperatures and maximum salinity. Furthermore, the same variables correlated to the  $F_{ST}$  matrix when comparing the Atlantic and Mediterranean clusters in addition to the range of temperature and mean salinity. The genetic distances obtained using the outliers with the Mediterranean localities correlated to minimum and range temperature.

The RDA including all the samples explained an accumulated 75.61% of the genetic variability found along its two first axes (Figure 5a). The first axis explained most of this variability and sorted the samples across a longitudinal gradient related mainly to maximum salinity, maximum temperature and mean salinity, with some incidence of the range of both environmental variables and the mean temperature. The second axis separated on one hand the two Atlantic populations and on the other hand the Adriatic



**TABLE 2** Comparative results of the analysis of candidate loci for selection

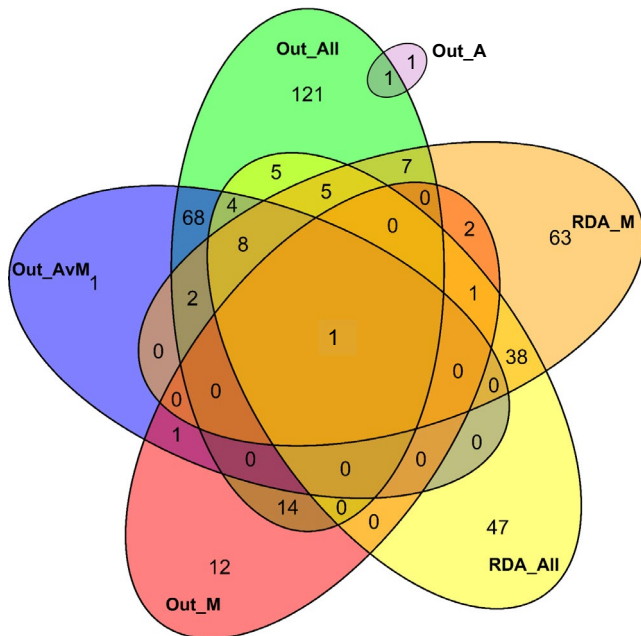
Analysis	Groupings	No. loci	Programs	Max $F_{ST}$	Mean $F_{ST} \pm SD$	Loci with blastn hits
Original dataset	Original dataset	3,348	NA	0.0863	$0.026 \pm 0.027$	NA
OAs	a) All localities	236	Arlequin + Lositan	0.3892	$0.088 \pm 0.093$	157 (59)
	b) Atlantic versus Mediterranean	85	Arlequin + Lositan	0.4911	$0.164 \pm 0.171$	54 (15)
	c) Mediterranean cluster	31	Arlequin + Lositan	0.2545	$0.085 \pm 0.074$	22 (9)
	d) Atlantic cluster	2	Lositan + Bayescan	0.2767	$0.068 \pm 0.073$	1 (0)
EAAs	a) All localities	109	Redundancy analysis	NA	NA	76 (26)
	b) Mediterranean	127	Redundancy analysis	NA	NA	88 (33)
Regional missing loci	NA	122	NA	NA	NA	33 (12)

Note: The first row shows the results of the original dataset with all loci for comparison. The other rows show the results of the analyses of candidate loci for selection using three different methodologies as stated in the first column: outlier analysis based on  $F_{ST}$  (OAs), environmental association analysis (EAAs) using RDA and regional missing loci. Different groupings were also used. For OAs: "All localities" includes all samples grouped by locality, "Atlantic versus Mediterranean" includes all samples grouped by their membership to the Atlantic or Mediterranean clusters, "Mediterranean cluster" includes all samples of the Mediterranean cluster grouped by locality, and "Atlantic cluster" includes all samples of the Atlantic cluster grouped by locality. The individuals in the Atlantic and Mediterranean clusters are those defined by the  $k$ -means clustering analysis with  $k = 2$ . For EAAs: "All localities" includes all localities and "Mediterranean" includes the Mediterranean localities excluding the transition zone. For each analysis, we provide the number of outlier loci identified, the programs used (the two most conservative in each case). For the OAs, we also provide the maximum and mean pairwise  $F_{ST}$ . The number of outlier loci with BLAST hits is given in the last column, and those to which a candidate function could be assigned are indicated in parentheses.

population of Trieste from the remaining Mediterranean populations. A total of 98 association loci were found, most of them related to maximum salinity, maximum temperature and mean temperature (Figure S8, Supplementary Dataset). The RDA including only the Mediterranean samples explained 43.96% of the variability found (Figure 5b), and the individuals were distributed longitudinally according to maximum temperature and mean and maximum salinity. The second axis separated the Adriatic population of Trieste and was explained by mean, minimum and range of temperature. A total of 127 association loci were found with the RDA using the Mediterranean samples, most of them related to maximum salinity and range of temperature (Figure S8, Supplementary Dataset). Some of the association loci found using RDA were also found as outlier loci using OAs, but most of them were different (Figure 3, Supplementary Dataset).

The analysis of regionally missing loci revealed that a total of 901 haplotype loci initially discarded during the filtering step (because they were not present in at least 70% of individuals) would have been retained if only samples from a particular area had been considered. Thus, missing loci were not randomly distributed across populations and regions. The frequency distribution of these regionally missing loci grouped the localities following the same structure found for the overall dataset (Figure 6). A total of 121 loci had a clear pattern of missingness: 45 loci were completely absent in the Atlantic while found in a high percentage of individuals of the other areas, and 76 loci were completely absent in the eastern Mediterranean while found at high percentages in individuals of the Atlantic and transition localities (Figure 6, Supplementary Dataset). None of these 121 loci was completely absent neither in the transition nor in the western Mediterranean areas.

Between 50% and 71% of the candidate loci identified with the different analysis (OAs and EAAs) and groupings had positive blastn hits ( $E$ -value below  $1 \times 10^{-4}$ ) to *Strongylocentrotus purpuratus* NCBI sequences, while 27% of the regional missing loci (Table 2). Overall, 305 loci had a blastn hit located on 233 different scaffolds, ranging from 1 to 8 loci per scaffold (Supplementary Dataset). A total of 184 scaffolds had only one single locus, while the remaining 49 scaffolds concentrated the other 121 loci. The 8 loci found in scaffold NW\_011967361.1 are separated by approximately 216Kb in *S. purpuratus*, correspond to exons and introns of six different genes and are found on different subsets of candidate loci. The 7 loci identified in scaffold NW\_011978358.1 are separated by c. 90Kb, correspond to exons from three genes and are found in the OAs overall and the Atlantic versus Mediterranean subsets and in the EAAs Atlantic and Mediterranean subsets. A total of 195 out of the 285 candidate loci found to be associated with a gene were located completely or partially in exonic regions. Around one third of the loci with a blastn hit had associated GO functions in all cases (Table 2). When considering all samples, GO functions were related mainly to apoptotic process, lysosomal transport and organization, response to oestrogen, protein ubiquitination and nucleobase-containing compound catabolism anatomical structure development among others (Figure S9a, Supplementary Dataset). When considering only the Mediterranean samples, similar functions appeared, such as apoptotic process, protein ubiquitination, anatomical structure development, while new functions appeared, such as vesicle-mediated transport or response to mechanical stimulus (Figure S9b). Regarding genes potentially preventing random mating, we found 1 gene coding for a receptor for egg jelly that was an outlier when considering either all populations or the Atlantic and



**FIGURE 3** Venn diagrams showing the number of candidate loci for adaptation identified using the different groupings and methodologies. The methods include “Out,” as the detection of outlier loci based on  $F_{ST}$  and “RDA” as the detection of association loci based on redundancy analysis. The groups include “All,” by using all samples grouped by locality; “AvM,” by using all samples grouped as belonging to either the Atlantic or Mediterranean clusters; “M” by all samples of the Mediterranean grouped by locality; and “A” by using all samples of the Atlantic grouped by locality

Mediterranean clusters and a sperm-associated antigen, found by RDA when considering all samples (Supplementary Dataset).

## 4 | DISCUSSION

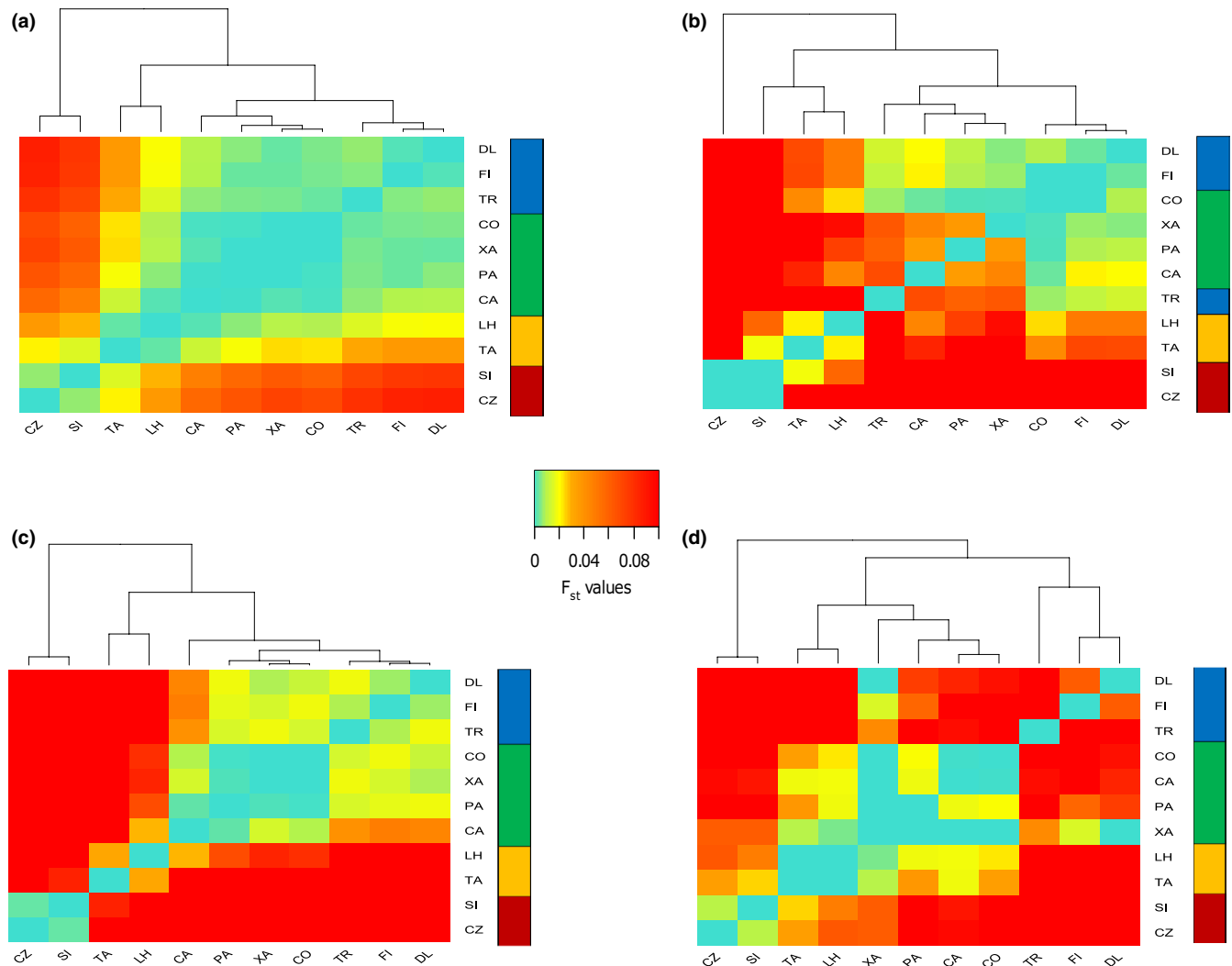
Using thousands of genome-wide markers, we detected so far unnoticed patterns of genetic structure among populations of *P. lividus* spanning most of its distribution range. We detected a gradient of differentiation following a longitudinal dimension, overlain by a major differentiation at the Atlantic–Mediterranean transition. Furthermore, specific analyses using candidate markers for adaptation enabled us to document population structure potentially driven by selection also at fine-scale level, using either OAs or EAAs. Different candidate markers for selection were detected when searched between and within subsets of genetically grouped individuals (Atlantic and Mediterranean clusters), likely indicating different selective pressures that would remain otherwise undetected without the hierarchical analysis approach followed in the present study, considering first all localities and subsequently analysing each identified genetic group. According to the genomic signals of association to environmental variables, adaptation to maximum salinity and temperature appeared as an important driver of the transition between Atlantic and Mediterranean basins, while temperature range associated signals were detected within the Mediterranean.

We also highlight how the use of the complete sequence information in the selected loci give a more detailed picture of genetic structure than using a single variable position (SNP) per locus.

### 4.1 | The Atlanto-Mediterranean transition

The major signal of genetic structuring detected in our data is the shift generated by the separation between Atlantic and Mediterranean populations. The existence of a genetic break associated with the oceanographic barrier to dispersion represented by the Almería–Oran front had been previously detected in *P. lividus* (Calderón et al., 2008; Duran et al., 2004) and is a common feature of many species spanning the Atlanto-Mediterranean transition (Pascual, Rives, Schunter, & Macpherson, 2017; Patarnello, Volckaert, & Castilho, 2007). However, our study unveils subtler patterns than previously found. The break between the Atlantic and the Mediterranean is not a clear-cut one. Rather, there is a transitional zone across the Gibraltar Strait through the Alboran Sea, as observed in other species (Pascual et al., 2016; Schunter et al., 2011). The Alboran Sea has a complex hydrology, with a system of gyres (Millot, 2005; Sánchez-Garrido, Lafuente, Fanjul, Sotillo, & De Los Santos, 2013) whose intensity and relative position can vary with time (Oguz, Macias, Garcia-Lafuente, Pascual, & Tintore, 2014). Considering the two identified genetic clusters (Atlantic and Mediterranean clusters), the transitional populations had a proportion of individuals of each cluster closely matching their geographic distance to the oceanographic break. Accordingly, the genetic diversity is higher in this transition zone because those localities are a mixture of the genetic pools found at both sides of the barrier. Furthermore, the Almería–Oran front is not a permanent barrier; rather, it is sensitive to fluctuations in circulation patterns that influence connectivity across oceans as found for the crab *Liocarcinus depurator* (Pascual et al., 2016). There is also evidence that temporal fluctuations in oceanographic currents may affect the genetic structure of the Mediterranean populations of *P. lividus* in the southern Iberian Peninsula (Calderón et al., 2012). The emerging scenario is thus that of a semi-permeable barrier to dispersion resulting in differentiation between oceans while at the same time allowing for a gradual mixture of the two genetic pools over a transitional area.

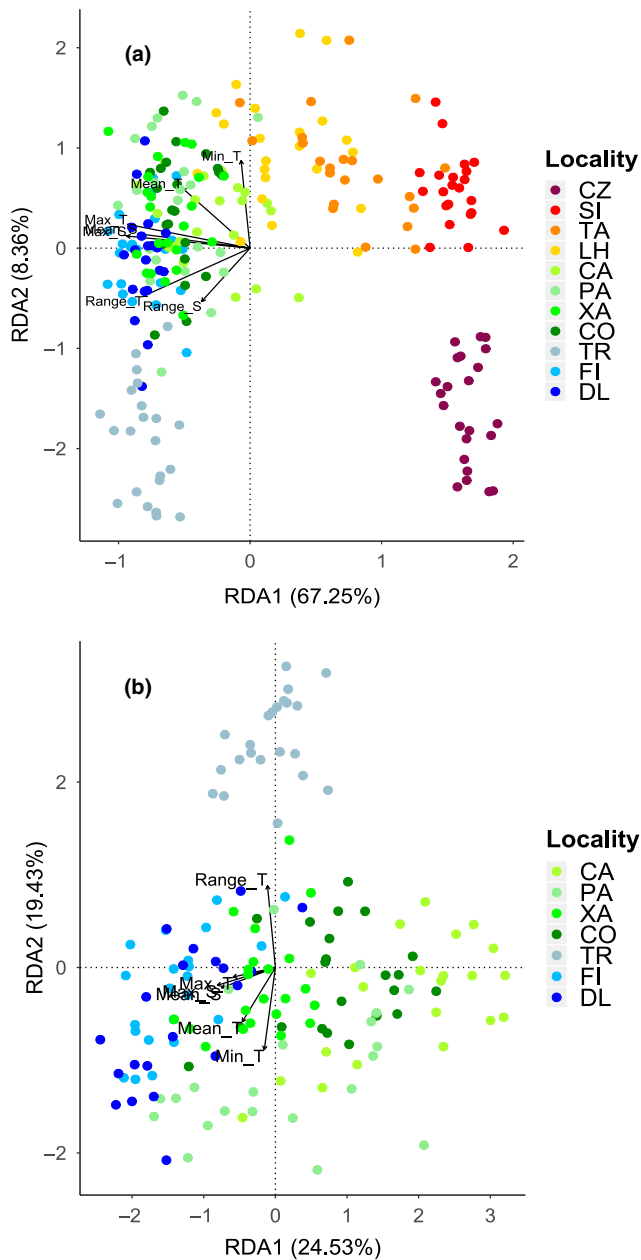
However, this population structure is likely to be maintained by contrasting selection pressures, as shown by our results. Thus, local adaptation can maintain genetic differentiation mediated by environmental heterogeneity despite the levels of gene flow occurring through the transition zone. Outlier loci obtained when considering the Atlantic versus Mediterranean clusters are almost all included in the subset obtained when considering all samples (Figure 3) which suggests that the Atlanto-Mediterranean break is also a major driver for the overall genetic structuring caused by selection. This is reflected by the fact that these two outlier subsets correlate with the same environmental variables, such as maximum temperature and salinity. The impact of these variables in the structuring of the whole dataset is also found in the analysis of association loci with



**FIGURE 4** Heatmaps and dendrograms based on  $F_{ST}$  pairwise distances among *Paracentrotus lividus* localities using four different sets of outlier loci obtained with OAs. (a) using all 3,348 loci, (b) using the 236 outlier loci obtained when grouping all samples by locality, (c) using the 85 outlier loci obtained when grouping all samples by its membership to either the Atlantic or Mediterranean clusters and (d) using the 31 outlier loci obtained when grouping only the samples of the Mediterranean cluster by locality. The colour bars on the right of the heatmaps show whether the localities belong to the Atlantic (red), the transition zone (orange), the western Mediterranean (green) or the eastern Mediterranean (blue) groups identified in the dendrograms. Note that the order of the populations may change according to the dendrogram construction. Sampling sites are as follows: CZ: Crozon; SI: Sidi Ifni; TA: Tarifa; LH: La Herradura; CA: Carboneras; PA: Palos; XA: Xàbia; CO: Colera; TR: Trieste, FI: Fier and DL: Dalyan

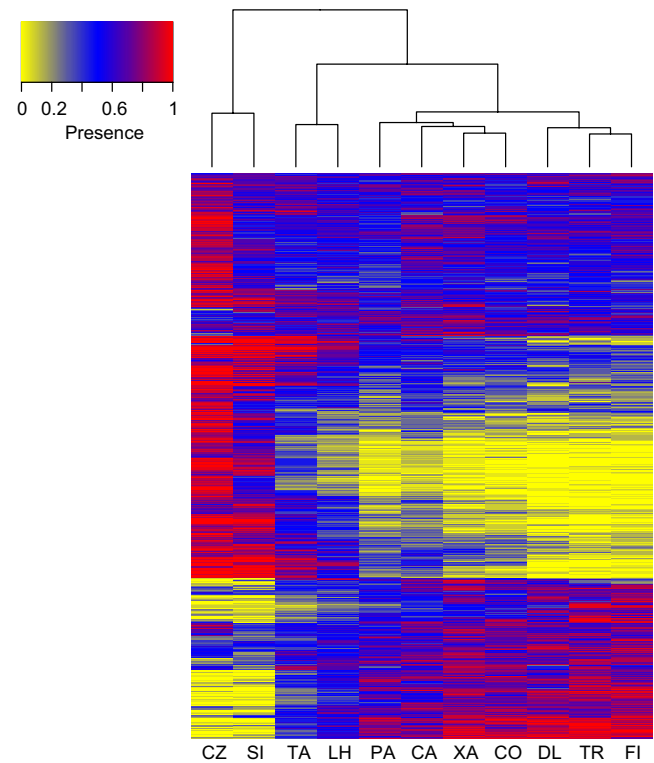
the RDA. Lysosomal transport and organization appeared as one of the major blocks of biological functions on the candidate loci in the comparison between Atlantic and Mediterranean with some functions linked to homeostasis such as in ion binding or transport. Ion transport has been frequently related to osmoregulation (Guh, Lin, & Hwang, 2015; Marshall & Grosell, 2005). Environmental values obtained from our sampling sites showed drastic changes from west to east, correlating with well-known longitudinal gradients of oceanographic parameters (UNEP/MAP, 2012). Even at small geographic scales, salinity values can change from c. 36.2‰ at the Atlantic side of the Gibraltar Strait to over 37.5‰ at the east of the Almería-Oran front (Brankart & Bresseur, 1998; Hidalgo et al., 2015; Pascual et al., 2016). These gradients effectively pose selective pressures on a group known to have limited abilities to regulate

their osmosis and ion concentrations (Binyon, 1972). Behavioural changes related to different salinities have been described in echinoderm species (Agüera, Schellekens, Jansen, & Smaal, 2015; Talbot & Lawrence, 2002) indicating that salinity may be a strong selective pressure for echinoderms. Interestingly, we found a locus under selection (in both the "all localities" outliers and the "Atlantic versus Mediterranean" outliers) that matched a receptor protein for egg jelly (REJ7, locus 1,361,733 in Supplementary Dataset). The REJ proteins are involved in sperm-egg binding (Gunaratne et al., 2007) and are thus potential sources of assortative mating dependent on compatible protein types and potential drivers for structuring or even speciation (Vilela-Silva, Hirohashi, & Mouro, 2008). For this reason, sequencing of some REJ proteins has been used to test for population structuring associated with egg-sperm recognition in sea



**FIGURE 5** Redundancy analysis plots. Redundancy analysis (RDA) performed with environmental data and SNP loci. RDA performed (a) with all the individuals and (b) with the individuals of the Mediterranean localities excluding the transition zone. Sampling sites are as follows: CZ: Crozon; SI: Sidi Ifni; TA: Tarifa; LH: La Herradura; CA: Carboneras; PA: Palos; XA: Xàbia; CO: Colera; TR: Trieste; FI: Fier and DL: Dalyan. Coloured points represent individual sea urchins coded by locality. Vectors indicate the environmental predictors of the two first RDA components

urchins (Muller, Heyden, Bowie, & Matthee, 2012), in a manner similar to that described for the *bindin* proteins (Calderón, Turon, Turon, & Lessios, 2009). We found a sperm-associated antigen when using all samples for the RDA associated to mean temperature, although no information has been found in the literature about the potential function of this gene. Further studies are necessary to understand the role of gamete recognition mechanisms in maintaining genetic



**FIGURE 6** Regionally missing loci. Heatmap and dendrogram of the percentage of individuals presenting the regionally missing loci that were discarded from the whole dataset but that were present on a specific area including the Atlantic (CZ, SI), the transition zone (TA, LH), the western Mediterranean (PA, CA, XA, CO) and the eastern Mediterranean (DL, TR, FI). Each colour represents the percentage of individuals within the locality that present valid genotypes for each locus, in rows. The dendrogram groups the localities by similarity of presence values of regionally missing loci

differentiation in *P. lividus* across its wide distribution. Our finding that 40% of the candidate loci with blastn hit were on scaffolds with multiple hits suggests that some genome regions are hotspots for selective pressures, something that could be better addressed when additional genomic tools, such as chromosome mapped reference genomes, become available.

The analysis of loci discarded because they exceeded the maximum missingness allowed in the pipeline also highlights the strength of the Atlanto-Mediterranean axis of genetic differentiation. The analysis of loci absent from a particular group of populations (likely as a result of mutations in the restriction targets) may carry information about particular characteristics of this group, including the drivers of their differentiation from the populations with these loci. In our case, the frequencies of presence of discarded loci showed a clear longitudinal pattern, with an accumulation of missing loci at both ends of the distribution of *P. lividus* (Figure 5). A generalized effect of genetic drift fixating null alleles at the edge of the distribution area, with unfavourable conditions resulting in small population sizes, seems unlikely as both the Atlantic and eastern Mediterranean populations have high densities of individuals (Boudouresque & Verlaque, 2001). Moreover, we have found similar genetic diversity

in all localities regardless of their situation in the distribution area indicating that Atlantic and eastern Mediterranean populations are not marginal and thus not more prone to genetic drift. The frequency patterns of these regional missing loci can be the result of historical processes, mutations appearing at a given locality and migrating towards neighbouring localities and further maintained by adaptive processes. The blast results pointed to several genes associated to ciliary movement, and once again to ion transport. As the same biological processes found in the candidate loci explaining the Atlanto-Mediterranean transition seem to also impact these regional missing loci, selection determining their frequency seems a plausible explanation.

## 4.2 | The Mediterranean internal structuring

Perhaps the most interesting results from the present study are the ones that become evident once the dominant axis of genetic structuring across the Atlanto-Mediterranean gradient is removed from the equation following a hierarchical approach. The ordination analysis of our data (Figure 2b) showed a clear arch or horseshoe effect (Diaconis, Goel, & Holmes, 2008), indicative of a latent west to east gradient in the data. This gradient goes from the Atlantic into the western Mediterranean and to the eastern Mediterranean, matching the longitudinal position of the populations. This result agrees with Maltagliati et al. (2010) who also found a progressive change in the frequencies of the three haplogroups defined in this species for the mitochondrial cytochrome *b* gene following a longitudinal dimension. This gradient is not an artefact of the Atlantic–Mediterranean transition, since when the Mediterranean cluster was analysed separately, the same longitudinal pattern appeared (Figure S3b). Mantel tests were also significant when including all localities or only the Mediterranean ones, indicating isolation by distance as a driver for differentiation in this species with high dispersal capabilities. However, even if the DAPC showed a continuum of populations, cluster analysis based on  $F_{ST}$  pairwise distances allowed us to define four groups (Atlantic, transition zone, W Mediterranean, E Mediterranean). Likewise,  $F_{ST}$  estimates showed that, in many cases, there is a significant differentiation among localities within each group, even the close ones (e.g. Trieste and Fier within the Adriatic Sea).

This within Mediterranean structuring also shows some within sea signals of the role of selection. The candidate loci found with either OAs and RDA when considering only the Mediterranean populations resulted on new sets of loci that are not driving the Atlanto-Mediterranean transition (Figure 3) and that showed a clearly different pattern of genetic variation (Figures 4d and 5b). The power of detecting outlier loci by OAs relies on the baseline  $F_{ST}$  defined by neutral markers (Ahrens et al., 2018). Thus, the presence of a strong barrier to dispersion, such as the Almería–Oran front, increases the maximum values of  $F_{ST}$  defined by the neutral differentiation and thus the threshold to define outliers (Figure S5). A similar effect could be found while looking for association loci with RDA, as local

environmental patterns may be obscured by general gradients when performing EAAs, and thus, the results are influenced by the scale and design of sampling (Ahrens et al., 2018). Consequently, the associated loci within the Mediterranean remained undetected until we analysed this region independently. Our results indicate that, in order to extract all information from the data, it is not enough to perform a candidate loci analysis considering all populations studied. Instead, a careful hierarchical analysis from the global, considering all localities, to the partitioned datasets, following genetic groups, allows a more in-depth study of the selective pressures acting between and within groups. In our case, the loci detected within the Mediterranean had BLAST hits that corresponded to functions of the cell cycle and DNA repair. These functions can be related to response to DNA damage (Bartek & Lukas, 2007; Ermolaeva & Schumacher, 2014). Which kind of stress induces these selective responses cannot be ascertained at present. In any case, it is remarkable that no lysosomal ion transport or organization functions were found, thus suggesting that within the Mediterranean, the role of salinity on the overall structuring might be much less important than when considering the whole distribution of the species, with temperature range emerging as more highly impacting population differentiation.

## 4.3 | Conservation and management implications

Population genomic results are key for conservation and management of ecologically relevant species. In the case of *P. lividus*, changes in the population dynamics can have a double-sided effect on benthic ecosystems. On one hand, it controls the abundance of macrophytes, so any increase in its densities can result in loss of habitat-forming algal species and generation of barren grounds (Agnetta et al., 2015; Sala et al., 1998). On the other hand, *P. lividus* at low densities can control opportunistic algal proliferation (Palacín et al., 1997; Bulleri, Benedetti-Cecchi, & Cinelli, 1999), generating patches for settlement of organisms and keeping the overall diversity. Thus, the density of this species should be finely tuned to keep an ecological balance, and human impacts are altering this balance in both senses. While overfishing of predators result in increased sea urchin densities in some areas, *P. lividus* populations face several threats in others. For instance, *P. lividus* is intensely harvested in many countries. Episodes of mortality have been attributed to global warming (Girard, Clemente, Toledo-Guedes, Brito, & Hernández, 2012; Yeruham et al., 2015), and the tropicalization of the Mediterranean may also be favouring the competing sea urchin *Arbacia lixula* to the detriment of *P. lividus* populations (Gianguzza et al., 2011; Pérez-Portela et al., 2019; Templado, 2014; Wangenstein, Dupont, Dupont, Casties, Turon, & Palacín, 2013; Wangenstein, Turon, Turon, Casso, & Palacín, 2013). Management of this key species requires knowledge of its genetic structure, connectivity and local adaptation. Population genomics has shown that genetic interchange occurs widely but has some limitations: the species is sensitive to dispersal barriers and displays isolation by distance. It has also been shown that a variety of selective pressures can be acting



on the genetic differentiation of the populations, salinity changes likely prominent among them but also temperature at regional level. In addition, although having a long PLD, there is not a panmictic population, and thus not a single genetic pool in the area studied. Consequently, then, *P. lividus* could be more vulnerable than previously thought, and current population declines in large areas (Rilov, 2016; Yeruham et al., 2015) should raise concern and promote management measures.

## ACKNOWLEDGEMENTS

We thank Emma Cebrian and Sandra Garcés for help with the specimens' collection, and Silvia Frias for help during DNA extractions. This work was supported by the Spanish Government projects ChallenGen CTM2013-48163 and PopCOMics CTM2017-88080 (MCIU, AEI/FEDER, UE). This is a contribution from the Consolidated Research Group "Benthic Biology and Ecology" SGR2017-1120 (Catalan Government).

## DATA AVAILABILITY STATEMENT

Raw read data from all individuals will be submitted to an NCBI SRA Bioproject upon acceptance. A list of genotypes of the haplotype loci dataset for all individuals, including sampling location, will be available in an ARLEQUIN file as a Supplementary Dataset upon acceptance.

## ORCID

Carlos Carreras  <https://orcid.org/0000-0002-2478-6445>

Marta Pascual  <https://orcid.org/0000-0002-6189-0612>

Xavier Turon  <https://orcid.org/0000-0002-9229-5541>

## REFERENCES

- Agnetta, D., Badalamenti, F., Ceccherelli, G., Di Trapani, F., Bonaviri, C., & Gianguzza, P. (2015). Role of two co-occurring Mediterranean sea urchins in the formation of barren from *Cystoseira* canopy. *Estuarine, Coastal and Shelf Science*, 152, 73–77. <https://doi.org/10.1016/j.ecss.2014.11.023>
- Agüera, A., Schellekens, T., Jansen, J. M., & Smaal, A. C. (2015). Effects of osmotic stress on predation behaviour of *Asterias rubens* L. *Journal of Sea Research*, 99, 9–16. <https://doi.org/10.1016/j.seares.2015.01.003>
- Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: Trends, biases and progress. *Molecular Ecology*, 27, 1342–1356. <https://doi.org/10.1111/mec.14549>
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a  $F_{st}$ -outlier method. *BMC Bioinformatics*, 9, 323. <https://doi.org/10.1186/1471-2105-9-323>
- Barnes, D. K. A., & Crook, A. C. (2001). Implications of temporal and spatial variability in *Paracentrotus lividus* populations to the associated commercial coastal fishery. *Hydrobiologia*, 465, 95–102.
- Bartek, J., & Lukas, J. (2007). DNA damage checkpoints: From initiation to recovery or adaptation. *Current Opinion in Cell Biology*, 19, 238–245. <https://doi.org/10.1016/j.ceb.2007.02.009>
- Benedetti-Cecchi, L., Bulleri, F., & Cinelli, F. (1998). Density dependent foraging of sea urchins in shallow subtidal reefs on the west coast of Italy (western Mediterranean). *Marine Ecology Progress Series*, 163, 203–211. <https://doi.org/10.3354/meps163203>
- Benestan, L., Ferchaud, A.-L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J. P., Baums, I. B., ... Luikart, G. (2016). Conservation genomics of natural and managed populations: Building a conceptual and practical framework. *Molecular Ecology*, 25, 2967–2977. <https://doi.org/10.1111/mec.13647>
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13), 3299–3315. <https://doi.org/10.1111/mec.13245>
- Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., ... Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, 25, 5073–5092. <https://doi.org/10.1111/mec.13811>
- Binyon, J. (1972). *Physiology of echinoderms*. Oxford; New York: Pergamon Press.
- Boudouresque, C. F., & Verlaque, M. (2001). Ecology of *Paracentrotus lividus*. In J. M. Lawrence (Ed.), *Edible sea urchins: Biology and ecology* (pp. 177–215). Tampa, FL: Elsevier.
- Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, 24, 5130–5144. <https://doi.org/10.1111/mec.13395>
- Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., ... Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, 277, 3725–3734. <https://doi.org/10.1098/rspb.2010.0985>
- Brankart, J. M., & Brasseur, P. (1998). The general circulation in the Mediterranean Sea: A climatological approach. *Journal of Marine Systems*, 18, 41–70. [https://doi.org/10.1016/s0924-7963\(98\)00005-0](https://doi.org/10.1016/s0924-7963(98)00005-0)
- Bulleri, F., Benedetti-Cecchi, L., & Cinelli, F. (1999). Grazing by the sea urchins *Arbacia lixula* L. and *Paracentrotus lividus* Lam. in the Northwest Mediterranean. *Journal of Experimental Marine Biology and Ecology*, 241, 81–95. [https://doi.org/10.1016/s0022-0981\(99\)00073-8](https://doi.org/10.1016/s0022-0981(99)00073-8)
- Calderón, I., Giribet, G., & Turon, X. (2008). Two markers and one history: Phylogeography of the edible common sea urchin *Paracentrotus lividus* in the Lusitanian region. *Marine Biology*, 154, 137–151. <https://doi.org/10.1007/s00227-008-0908-0>
- Calderón, I., Palacín, C., & Turon, X. (2009). Microsatellite markers reveal shallow genetic differentiation between cohorts of the common sea urchin *Paracentrotus lividus* (Lamarck) in northwest Mediterranean. *Molecular Ecology*, 18, 3036–3049. <https://doi.org/10.1111/j.1365-294x.2009.04239.x>
- Calderón, I., Pita, L., Brusciotti, S., Palacín, C., & Turon, X. (2012). Time and space: Genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Marine Biology*, 159, 187–197. <https://doi.org/10.1007/s00227-011-1799-z>
- Calderón, I., Turon, X., & Lessios, H. A. (2009). Characterization of the sperm molecule bindin in the sea urchin genus *Paracentrotus*. *Journal of Molecular Evolution*, 68, 366–376. <https://doi.org/10.1007/s00239-009-9219-4>
- Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation. *Scientific Reports*, 7, 43417. <https://doi.org/10.1038/srep43417>
- Carreras-Carbonell, J., Macpherson, E., & Pascual, M. (2006). Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, 15, 3527–3539. <https://doi.org/10.1111/j.1365-294x.2006.03003.x>

- Casso, M., Turon, X., & Pascual, M. (2019). Single zooids, multiple loci: Independent colonisations revealed by population genomics of a global invader. *Biological Invasions*, 21, 3575–3592. <https://doi.org/10.1007/s10530-019-02069-8>
- Couvray, S., & Coupé, S. (2018). Three-year monitoring of genetic diversity reveals a micro-connectivity pattern and local recruitment in the broadcast marine species *Paracentrotus lividus*. *Heredity (Edinb.)*, 120, 110–124. <https://doi.org/10.1038/s41437-017-0013-6>
- Diaconis, P., Goel, S., & Holmes, S. (2008). Horseshoes in multidimensional scaling and local kernel methods. *The Annals of Applied Statistics*, 2, 777–807. <https://doi.org/10.1214/08-aos165>
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14, 927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- Duran, S., Palacin, C., Becerro, M. A., Turon, X., & Giribet, G. (2004). Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). *Molecular Ecology*, 13, 3317–3328. <https://doi.org/10.1111/j.1365-294X.2004.02338.x>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A web-site and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Ermolaeva, M. A., & Schumacher, B. (2014). Systemic DNA damage responses: Organismal adaptations to genome instability. *Trends in Genetics*, 30, 95–102. <https://doi.org/10.1016/j.tig.2013.12.001>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27(9), 2215–2233. <https://doi.org/10.1111/mec.14584>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27, 489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Gianguzza, P., Agnetta, D., Bonaviri, C., Di Trapani, F., Visconti, G., Gianguzza, F., & Riggio, S. (2011). The rise of thermophilic sea urchins and the expansion of barren grounds in the Mediterranean Sea. *Chemical Ecology*, 27, 129–134. <https://doi.org/10.1080/02757540.2010.547484>
- Girard, D., Clemente, S., Toledo-Guedes, K., Brito, A., & Hernández, J. C. (2012). A mass mortality of subtropical intertidal populations of the sea urchin *Paracentrotus lividus*: Analysis of potential links with environmental conditions. *Marine Ecology*, 33, 377–385. <https://doi.org/10.1111/j.1439-0485.2011.00491.x>
- Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36, 3420–3435. <https://doi.org/10.1093/nar/gkn176>
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Guh, Y.-J., Lin, C.-H., & Hwang, P.-P. (2015). Osmoregulation in zebrafish: Ion transport mechanisms and functional regulation. *EXCLI Journal*, 14, 627–659.
- Gunaratne, H. J., Moy, G. W., Kinukawa, M., Miyata, S., Mah, S. A., & Vacquier, V. D. (2007). The 10 sea urchin receptor for egg jelly proteins (SpREJ) are members of the polycystic kidney disease-1 (PKD1) family. *BMC Genomics*, 8, 235. <https://doi.org/10.1186/1471-2164-8-235>
- Hapke, A., & Thiele, D. (2016). GlibPs: A toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular Ecology Resources*, 16, 979–990. <https://doi.org/10.1111/1755-0998.12510>
- Hereu, B., Zabala, M., Linares, C., & Sala, E. (2004). Temporal and spatial variability in settlement of the sea urchin *Paracentrotus lividus* in the NW Mediterranean. *Marine Biology*, 144, 1011–1018. <https://doi.org/10.1007/s00227-003-1266-6>
- Hereu, B., Zabala, M., Linares, C., & Sala, E. (2005). The effects of predator abundance and habitat structural complexity on survival of juvenile sea urchins. *Marine Biology*, 146, 293–299. <https://doi.org/10.1007/s00227-004-1439-y>
- Hidalgo, M., Reglero, P., Álvarez-Berastegui, D., Torres, A. P., Álvarez, I., Rodríguez, J. M., ... Alemany, F. (2015). Hidden persistence of salinity and productivity gradients shaping pelagic diversity in highly dynamic marine ecosystems. *Marine Environment Research*, 104, 47–50. <https://doi.org/10.1016/j.marenvres.2015.01.001>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., ... Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, 109, 19345–19350. <https://doi.org/10.1073/pnas.1216128109>
- Maltagliati, F., Di Giuseppe, G., Barbieri, M., Castelli, A., & Dini, F. (2010). Phylogeography and genetic structure of the edible sea urchin *Paracentrotus lividus* (Echinodermata: Echinoidea) inferred from the mitochondrial cytochrome b gene. *Biological Journal of the Linnean Society*, 100, 910–923. <https://doi.org/10.1111/j.1095-8312.2010.01482.x>
- Marshall, W. S., & Grosell, M. (2005). Ion transport, osmoregulation, and acid-base balance. In D. Evans & J. Claiborne (Eds.), *Physiology of fishes* (Vol. 3, pp. 177–230). Boca Raton, FL: CRC Press.
- Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., ... Bargelloni, L.; FishPopTrace Consortium. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23, 118–135. <https://doi.org/10.1111/mec.12568>
- Millot, C. (2005). Circulation in the Mediterranean Sea: Evidences, debates and unanswered questions. *Scientia Marina*, 69, 5–21. <https://doi.org/10.3989/scimar.2005.69s15>
- Minchin, D. (1992). Multiple species, mass spawning events in an Irish sea lough: The effect of temperatures on spawning and recruitment of invertebrates. *Invertebrate Reproduction & Development*, 22, 229–238. <https://doi.org/10.1080/07924259.1992.9672275>
- Muller, C. M., Von Der Heyden, S., Bowie, R. C. K., & Matthee, C. A. (2012). Oceanic circulation, local upwelling and palaeoclimatic changes linked to the phylogeography of the Cape sea urchin

- Parechinus angulosus*. *Marine Ecology Progress Series*, 468, 203–215. <https://doi.org/10.3354/meps09956>
- Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.
- Oguz, T., Macias, D., Garcia-Lafuente, J., Pascual, A., & Tintore, J. (2014). Fueling plankton production by a meandering frontal jet: A case study for the Alboran Sea (Western Mediterranean). *PLoS ONE*, 9, e111482. <https://doi.org/10.1371/journal.pone.0111482>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., ... Wagner, H. (2019). Package “vegan”. Community Ecology Package.
- Palacín, C., Giribet, G., Garner, S., Dantart, L., & Turon, X. (1998). Low densities of sea urchins influence the structure of algal assemblages in the western Mediterranean. *Journal of Sea Research*, 39, 281–290. [https://doi.org/10.1016/S1385-1101\(97\)00061-0](https://doi.org/10.1016/S1385-1101(97)00061-0)
- Palacín, C., Giribet, G., & Turon, X. (1997). Patch recolonization through migration by the echinoid *Paracentrotus lividus* in communities with high algal cover and low echinoid densities. *Cahiers De Biologie Marine*, 38, 267–271.
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Pascual, M., Palero, F., Hugo Garcia-Merchan, V., Macpherson, E., Robainas-Barcia, A., Mestres, F., ... Abello, P. (2016). Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. *Scientific Reports*, 6(1), <https://doi.org/10.1038/srep29892>
- Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS ONE*, 12(5), e0176419. <https://doi.org/10.1371/journal.pone.0176419>
- Patarnello, T., Volckaert, F., & Castilho, R. (2007). Pillars of hercules: Is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, 16, 4426–4444. <https://doi.org/10.1111/j.1365-294X.2007.03477.x>
- Paterno, M., Schiavina, M., Aglieri, G., Souissi, J. B., Boscarì, E., Casagrandi, R., ... Melià, P. (2017). Population genomics meet Lagrangian simulations: Oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* in the Adriatic and Ionian seas. *Ecology and Evolution*, 7(8), 2463–2479.
- Pedrotti, M. L. (1993). Spatial and temporal distribution and recruitment of Echinoderm larvae in the Ligurian Sea. *Journal of the Marine Biological Association of the United Kingdom*, 73, 513–530. <https://doi.org/10.1017/S0025315400033075>
- Penant, G., Aurelle, D., Feral, J.-P., & Chenuil, A. (2013). Planktonic larvae do not ensure gene flow in the edible sea urchin *Paracentrotus lividus*. *Marine Ecology Progress Series*, 480, 155–170. <https://doi.org/10.3354/meps10194>
- Pérez-Portela, R., Wangenstein, O. S., Garcia-Cisneros, A., Valero-Jiménez, C., Palacín, C., & Turon, X. (2019). Spatio-temporal patterns of genetic variation in *Arbacia lixula*, a thermophilous sea urchin in expansion in the Mediterranean. *Heredity (Edinb.)*, 122, 244–259. <https://doi.org/10.1038/s41437-018-0098-6>
- Pierce, D. (2017). Package “ncdf4” interface to Unidata netCDF (Version 4 or Earlier). Format Data Files.
- Privitera, D., Noli, M., Falugi, C., & Chiantore, M. (2011). Benthic assemblages and temperature effects on *Paracentrotus lividus* and *Arbacia lixula* larvae and settlement. *Journal of Experimental Marine Biology and Ecology*, 407, 6–11. <https://doi.org/10.1016/j.jembe.2011.06.030>
- Rilov, G. (2016). Multi-species collapses at the warm edge of a warming sea. *Scientific Reports*, 6, 36897. <https://doi.org/10.1038/srep36897>
- Ryman, N., Palm, S., André, C., Carvalho, G. R., Dahlgren, T. G., Jorde, P. E., ... Ruzzante, D. E. (2006). Power for detecting genetic divergence: Differences between statistical methods and marker loci. *Molecular Ecology*, 15, 2031–2045. <https://doi.org/10.1111/j.1365-294X.2006.02839.x>
- Sala, E., Boudouresque, C. F., & Harmelin-Vivien, M. (1998). Fishing, trophic cascades, and the structure of algal assemblages: Evaluation of an old but untested paradigm. *Oikos*, 82(3), 142–153. <https://doi.org/10.2307/3546364>
- Sánchez-Garrido, J. C., Lafuente, J. G., Fanjul, E. Á., Sotillo, M. G., & De Los Santos, F. J. (2013). What does cause the collapse of the Western Alboran Gyre? Results of an operational ocean model. *Progress in Oceanography*, 116, 142–153. <https://doi.org/10.1016/j.pocean.2013.07.002>
- Sandoval-Castillo, J., Robinson, N. A., Hart, A. M., Strain, L. W. S., & Beheregaray, L. B. (2018). Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Molecular Ecology*, 27, 1603–1620. <https://doi.org/10.1111/mec.14526>
- Schunter, C., Carreras-Carbonell, J., MacPherson, E., Tintore, J., Vidal-Vijande, E., Pascual, A., ... Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, 20, 5167–5181. <https://doi.org/10.1111/j.1365-294X.2011.05355.x>
- Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS ONE*, 6(7), e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Talbot, T. D., & Lawrence, J. M. (2002). The effect of salinity on respiration, excretion, regeneration and production in *Ophiophragmus filigraneus* (Echinodermata: Ophiuroidea). *Journal of Experimental Marine Biology and Ecology*, 275, 1–14. [https://doi.org/10.1016/S0022-0981\(02\)00208-3](https://doi.org/10.1016/S0022-0981(02)00208-3)
- Templado, J. (2014). Future trends of mediterranean biodiversity. In S. Goffredo & Z. Dubinsky (Eds.), *The Mediterranean Sea* (pp. 479–498). Dordrecht, The Netherlands: Springer.
- Tomas, F., Romero, J., & Turon, X. (2004). Settlement and recruitment of the sea urchin *Paracentrotus lividus* in two contrasting habitats in the Mediterranean. *Marine Ecology Progress Series*, 282, 173–184. <https://doi.org/10.3354/meps282173>
- UNEP/MAP UNEP. (2012). *State of the Mediterranean marine and coastal environment*. Athens: Barcelona Convention.
- Vandamme, S. G., Maes, G. E., Raeymaekers, J. A. M., Cottenie, K., Imsland, A. K., Hellemans, B., ... Volckaert, F. A. M. (2014). Regional environmental pressure influences population differentiation in turbot (*Scophthalmus maximus*). *Molecular Ecology*, 23, 618–636. <https://doi.org/10.1111/mec.12628>
- Vilela-Silva, A.-C.-E.-S., Hirohashi, N., & Mouro, P. A. S. (2008). The structure of sulfated polysaccharides ensures a carbohydrate-based mechanism for species recognition during sea urchin fertilization. *International Journal of Developmental Biology*, 52, 551–559. <https://doi.org/10.1387/ijdb.072531av>
- Wangenstein, O. S., Dupont, S., Casties, I., Turon, X., & Palacín, C. (2013). Some like it hot: Temperature and pH modulate larval development and settlement of the sea urchin *Arbacia lixula*. *Journal of Experimental Marine Biology and Ecology*, 449, 304–311. <https://doi.org/10.1016/j.jembe.2013.10.007>
- Wangenstein, O. S., Turon, X., Casso, M., & Palacín, C. (2013). The reproductive cycle of the sea urchin *Arbacia lixula* in northwest Mediterranean: Potential influence of temperature and photoperiod. *Marine Biology*, 3, 3157–3168. <https://doi.org/10.1007/s00227-013-2303-8>
- Wangenstein, O. S., Turon, X., Garcia-Cisneros, A., Recasens, M., Romero, J., & Palacín, C. (2011). A wolf in sheep's clothing: Carnivory in dominant sea urchins in the Mediterranean. *Marine Ecology Progress Series*, 441, 117–128. <https://doi.org/10.3354/meps09359>
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, W., Schwartz, M.,

- et al. (2016). gplots: Various R programming tools for plotting data. Retrieved from <http://CRAN.R-project.org/package=gplots>
- Warnes, G., Gajanc, G., Leisch, F., & Man, M. (2019). Package "genetics". Population Genetics.
- White, T., van der Ende, J., & Nichols, T. E. (2019). Beyond Bonferroni revisited: Concerns over inflated false positive research findings in the fields of conservation genetics, biology, and medicine. *Conservation Genetics*, 20(4), 927–937. <https://doi.org/10.1007/s10592-019-01178-0>
- Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M.-J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, 27, 2347–2364. <https://doi.org/10.1111/mec.14589>
- Yeruham, E., Rilov, G., Shpigel, M., & Abelson, A. (2015). Collapse of the echinoid *Paracentrotus lividus* populations in the Eastern Mediterranean – Result of climate change? *Scientific Reports*, 5, 13479. <https://doi.org/10.1038/srep13479>

#### BIOSKETCH

The research of the authors' team focuses on the application of genetic tools to general and applied aspects of marine biodiversity assessment and conservation. We use population genetics and genomics to analyse the genetic structure of a variety of invertebrate and fish species. We focus on keystone and commercial species, as well as on introduced and invasive ones. We are also developing metabarcoding tools for biodiversity assessment and monitoring of natural benthic communities.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Carreras C, García-Cisneros A, Wangenstein OS, et al. East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Divers Distrib.* 2020;26:382–398. <https://doi.org/10.1111/ddi.13016>